Daniel Müllensiefen, Geraint Wiggins, David Lewis

High-level feature descriptors and corpus-based musicology: Techniques for modelling music cognition

1 Introduction

In recent years large electronic collections of music in a symbolically-encoded form have been made available. They have enabled music researchers to develop and test precise empirical theories of music on large data sets. Both the availability of music data and the development of new empirical theories creates a new perspective for Systematic Musicology, which, as a discipline, often sets out to explain or describe music through the induction of empirical laws, regularities or statistical correlations in relation to music objects or music related behaviour (see e.g. Karbusicky, 1979; Karbusicky & Schneider, 1980; Schneider, 1993; Huron, 1999; Parncutt, 2007). We present two methodological frameworks, feature-extraction and corpus-based musicology, which are the core approaches of a particular research project, M^{4}S, whose aim is to discover mechanisms of music cognition. These two frameworks are also very useful for many other empirical tasks in Systematic Musicology.

Before we go on to sketch a current project and to describe feature-extraction and corpus-based musicology, we need to clarify the term *symbolically-encoded music,* better to position our approach in the larger picture of contemporary music research. By "symbolically-encoded music", we mean music in a computer-readable format where the fundamental unit of representation is the note. Thus, symbols in these formats designate notes as might be performed by musicians, or played back by music software, or rendered to score notation by music engraving/publishing software. Well-known and widely-used symbolic computer formats include the Plaine and Easie Code (Brook, 1970), MIDI (International MIDI Association, 1988), kern** (Huron, 1995), MusicXML (Recordare, 2004/2007), and MuseData (Hewlett, 1997), Selfridge-Field (1997) gives a summary up to 1997. Symbolic formats can be contrasted with audio formats which, instead of capturing notes explicitly, encode the sonic aspect of a musical performance by representing sound as a complex waveform. The best known formats are audio CD, the WAV and AIFF formats used primarily in computers and iPods, and MPEG-1 Audio-Layer 3 (mp3) as a compression format used for web-based and portable applications.

The decision to use symbolic formats for our study lies in the fact that we are interested in objects of music cognition like melodies, rhythms, and harmonies, which seem to be mentally represented in a form comparable with symbolic encoding formats. Accordingly, our project will make use of experimental procedures that require the comparison, recognition or reproduction of pitch or rhythm sequences which also can be described most easily at a note level. To extract faithful representations of (monophonic) pitch or rhythm sequences from a polyphonic piece of music encoded in an audio format remains an unsolved problem. The procedures of instrument recognition, voice separation, and note segmentation necessary for our purposes currently suffer from unacceptably high error rates, especially when these procedures are combined to produce transcription from complex audio signals. But when we turn to symbolic formats the extraction of melodies, harmonies and rhythms as well as the extraction of features of these musical entities from polyphonic music becomes much easier, as we shall see below.

Context and goals of the M^{4}S project

The project Modelling Music Memory and the Perception of Melodic Similarity (M^{4}S) is a three-year project hosted in the Computing Department of Goldsmiths, University of London and aims to construct cognitive models of memory and of similarity perception for melodies, questions which fall in the areas of music psychology and music cognition. The way to arrive at these answers involves music analysis based on algorithms and the computation of distributions of analytic features. Therefore, scientific outcomes of the project can be found in the precise definition and evaluation of analytic features as well as in the identification patterns of musical features that are interesting for music analytic purposes (see section "Feature extraction").

The project is organised in four steps.

The first step involves the construction of a database for a large collection of symbolically-encoded music. This includes not only the incorporation of the raw musical data, but also the computation and integration of higher level musical structure (based on the concept of constituents, Smaill et al. 1993). Constituents denote musically meaningful groups, and are linked to sets of note events, allowing the annotation of a musical surface with structures and relationships.

The second step is an empirical evaluation of the defined features based on psychological experiments. As the number of potential implementations for any given feature is large, the goal of this step is to select the implementation of a feature that matches human perception most closely. Feature computation and evaluation will be explained in detail in the next section.

As a third step, the joined distributions of the selected features is computed. This includes the detection of associations or correlations between categories of different features which might be related to musical patterns. The distributions of feature categories and feature category combinations as they appear in the music corpus are recorded.

In a fourth and final step, these distributions are employed to model the data from a second series of psychological experiments. This time these experiments are focused on memory for music objects like melodies and rhythms.

The music corpus used for M^{4}S is a collection of 14,067 transcriptions of pop songs from the 1950s to 2006 in the MIDI file format. These files were commercially acquired and were produced, to a high standard, by musicians for a distributing company (Geerdes MIDI Music; http://www.midimusic.de) which normally sells its MIDI files to studio musicians, entertainers or karaoke enterprises where they are used as backing tracks. The MIDI files may be considered accurate transcriptions of the original recordings and contain all the vocal and instrumental voices that are present in the original. In addition, the encoding conventions of the distributor allow us automatically to identify the lyrics, main tune (vocal), bass line, drums, and harmony instruments. Apart from these conventions, the files are not otherwise annotated: there is no indication of song sections, harmonic labels, or abstracted representations of rhythmic and harmonic patterns. In short, the raw data that the M^4S project has to deal with are effectively[1] faithful full-score transcriptions of pop songs.

Whilst MIDI is not a standard designed for representing scores (it was originally intended for real-time synthesizer control), and lacks even such basic information as enharmonic pitch spelling, it is a remarkably rich format. A MIDI file consists of a sequence of events indicating, in the case of notes, pitch, start and finish times, mix volume, timbral strength ('key velocity'), instrumental timbre, voice, stereo placement, pitch bend, etc. Lyrics, key and time signatures, and transposing instruments are also supported.

In M^4S, all this information is recorded in an indexed SQL database, allowing not only fast querying, but also the construction of compound (or higher-level) musical queries.

Musical objects provide information about sets of other musical objects. For example, a melody is a set of notes which may be considered as musically meaningful and distinct, to some extent, from their surroundings. A melodic contour, however, might refer to that phrase and indicate that it has, say, an arch structure, or a sentence object might group several musical phrases rather as the melody groups several notes (for an example of the power of reasoning with such constructs, see Smaill et al, 1993).

Whereas the lowest level of data in our database is specific to the particular encoding format and conventions of the collection, these higher-level constructs are progressively more abstract, allowing musical objects to be directly compared even if their source material originates from different corpora, represented with different encodings and even different notations—potentially, even audio representations may be treated similarly. Musical objects are defined in terms of how they behave and what information they contain, rather than by the details of their implementation, in a practice known as Abstract Data Typing, allowing the commonalities between representations to be emphasized and exploited, without undermining the integrity of genuinely different sources and types of information. This allows us to run our algorithms on diverse corpora without modification, whilst at the same time permitting the details (parameters) of how those algorithms are carried out to vary between corpora. This, in turn, permits us to draw inference about how music from different repertoires might be similar or different with respect to analytic features, even if the repertoires stem from completely different technological sources. Apart from the pop music collection which was imported from MIDI files, the database system also hosts other repertoires, including the Essen folksong collection (Schaffrath, 1995), a collection of Chorale melodies from J.S. Bach's cantatas and a set of Canadian folksongs which were all imported from kern** source files.

When we refer to features and feature extraction, we refer to these musical objects. To be more specific, any musical object that we view as being meaningful or useful in itself (as opposed to being an intermediate step towards a useful end) can be described as a

[1] "Effectively" because some musical constructions which would be denoted symbolically in a score, such as crescendi and tempo changes, are encoded as parameter changes in the MIDI file.

feature. Examples of features are the sequence of notes in a melodic phrase, a summary of the melodic arch, the prevailing harmony over a period of time, or a formal label.

Our 14,067 songs were selected from the full catalogue of the distributor in such a way as to be as representative as possible of the history of commercial western pop music. Most of the songs were chart-listed or recorded by commercially successful artists. While songs from Anglo-American artists represent about 70% of the corpus, the remaining 30% are distributed primarily across other European countries. Stylistically, the corpus comprises everything from Rock'n'Roll to Soul and Funk classics to modern day HipHop and Ragga. The reason for choosing a corpus of highly commercial pop songs lies in the fact that this musical repertoire can be assumed to be very familiar to the musically untrained Western listeners from which the primary group of participants in the project's experiments will be drawn.

The choice of this specific music corpus corresponds with the hypothesis at the heart of the modelling idea behind M^4S. The hypothesis is that cognitive processes like memory encoding and retrieval as well as similarity perception are influenced by the familiarity of the musical material that is to be processed. As it is impossible to determine directly and objectively the familiarity of a particular subject with a musical object/stimulus, a viable approach is to approximate the individual familiarity with the frequency that a musical object has been listened to or, in a further approximation, with its frequency of occurrence in a corpus of music that the subject is highly familiar with. This hypothesis makes M^4S also a project about implicit statistical learning of non-verbal material.

Thus, the project has an explicit cognitive focus which distinguishes it from previous research projects that dealt with large music corpora. Perhaps the most ambitious corpus-based musicology project was one based in Princeton University concerned with Josquin scholarship. From 1963 to the beginning of the eighties, researchers, led by Arthur Mendel and Lewis Lockwood, generated electronic scholarly editions of the complete works of Josquin (as defined by the research of the time), many including concordances, and relevant related works. From this, statistics for cadential progressions and modal indicators were compiled and subjected to statistical analysis primarily in order to study issues of authorship and stemmatic filiation (see, for example, Lockwood, 1970 and various papers in *Computers in the Humanities* between 1969 and 1978). The ambitions of this project, though great, never extended to revealing cognitive processes, being limited, essentially, to style analysis.

In folk music research, feature extraction and the use of computers have been employed as a means for the (automatic) classification of songs (mainly melodies) according to their musical characteristics. In a comprehensive study Steinbeck (1982) classified European folk melodies into six homogeneous groups by employing Ward's classification algorithm with 35 relatively simple features derived from the monophonic melodies. He was able to show that this classification was in close correspondence with the melodies' regional origin and functional uses. Although theoretically proposed in his concepts of "Determination" and "Gängigkeit" (prevalence or commonness), Steinbeck did not make any use of the frequencies of occurrence of his features for computing similarities or grouping pieces. Starting from a similar background and music repertoire, Damien Sagrillo (1999) carried out an exploration of frequently employed melodic formulae in folk songs from Luxembourg and Lorraine. Among the outcomes of

Sagrillo's work is the ordering and classification of a catalogue of more than 3,000 melodic phrases and the description and discussion of frequent melodic patterns in this particular folksong repertoire.

In contrast, David Huron in his work from the last two decades (e.g. 1988, 2006) has paid special attention to the frequencies of features over a range of different music corpora. The features Huron has been using can be characterised as mainly simple features derived from monophonic melodies, like interval direction and interval size or melodic phrase contour. As summarised in his latest book (Huron, 2006), he and his collaborators have also tested the relationship between the frequencies of features and cognitive behaviour experimentally. In many instances, Huron was able to show a strong relation between feature frequencies and cognitive expectation and, thus, considers statistical learning as the major mechanism for the creation of knowledge about music. Going one step beyond Huron, M^4S tries to better approximate the implicit knowledge that people have about a specific musical style by using a music corpus that should be representative of the experimental participants' musical experience. Secondly, M^4S tries to work with richer music representations than the pre-segmented monodies that Huron uses in many of his studies and to employ more complex features which might more accurately reflect the music percepts of untrained listeners. We call these features "higher-level features" and examples are given in the feature section below.

2 Corpus-based musicology

The ideas behind the concept of *corpus-based musicology* are not entirely new, and have been explored by some of the projects briefly discussed above (Lockwood, 1970; Steinbeck, 1982; Sagrillo, 1999; Huron, 2006). Corpus-based musicology might be regarded as one concrete instantiation of empirical musicology as envisioned by the contributors of the popular volume by the same name (Clarke & Cook, 2004). Nonetheless, by introducing the novel term *corpus-based musicology*, we would like to emphasize the motivations that drive the musicological aspects of the M^4S project. They can be summarized as follows:

1. Analytical observations about a musical object as they are made by automatic feature extraction or human analysis are only really meaningful in context of a music corpus. When experienced music analysts observe peculiarities about a musical object a reference corpus serves always as the basis for comparison– though it might not be explicitly named. Indeed, the reason why analytic works by experts in their respective fields are generally preferred is that they have a wealth of knowledge about a particular musical style at their disposal; one may think of Charles Rosen's Classical Style (1976) or Allan Moore's (1997) monograph on Sgt. Pepper's Lonely Hearts Club Band, for example. When we trust the computer to analyse music and to extract and compare feature value then we have to make a conscious effort to introduce the context of a music corpus in which feature comparison is meaningful.
2. Using a computer for music-analytic purposes forces a researcher to define very precisely what are the analytic markers, features and structures that she or he wants to observe. For example, it might seem perfectly clear to every knowledgeable analyst what a melodic phrase is, but if we attempt to design an

algorithm that is supposed to indicate phrase boundaries for every unsegmented stream of notes in a corpus of songs, the definition of a melodic phrase becomes a surprisingly hard problem. One has to know in advance what melodic phenomena might occur in the corpus and how they are expressed in the music data representation. Decisions must then be made about how to treat certain classes of phenomena in terms of the desired outcome (i.e. to assign a number of boundaries in reasonable locations for each monophonic melody). In designing these algorithms, the researcher creates his/her own tools for observation and it is necessary that the tools are designed with respect to the peculiarities of the music corpus. To give a simple example: one could think of a phrase segmentation algorithm that uses a completely different set of criteria and procedures according to whether it is employed for the segmentation of vocal tunes from popular music or repetitive bass lines from 1970s funk pieces. To summarise: knowledge of a reference corpus necessarily informs the choice and design of the analytical tools that are used to explore it. This circular relationship between theory, tools, and data is only rarely fully acknowledged when feature extraction and analysis algorithms are made available (e.g. the MIDI tool box, Eerola & Toiviainen, 2004, or Cory McKay's jSymbolic feature extractor, McKay, 2006).

3. One of the main motivations and potentially most useful outcomes that can arise from corpus-based analysis is the *quantification* of musical structures in a music corpus. Knowledge about the frequency of occurrence of musical objects or their combinations can not only serve as a backbone in models of music cognition but might also be very valuable for the design of music information retrieval systems for that corpus. In addition, information about the frequency of musical structures in combination with meta-data e.g. about the time and geographical origin of the musical works, their genre categories or their popularity in specific eras creates a new perspective on music analysis: quantitative information about musical structures can then be associated with categories of styles and popularity. The analysis of the usage of melodic and harmonic formulae over different times and styles that Allan Moore did paradigmatically for the "doo-wop progression" (Moore, 2006) could be extended with the help of the computer for all significant harmonic formulae as well as for melodic and rhythmic patterns that are frequently used in a similar fashion. This could make an important contribution to the study of evolution of musical style, at least in many areas of popular music.

Such an approach can offer two advantages. Firstly, it can facilitate the discussion of common factors and trends within a corpus. Secondly, it can reveal aspects of a work or group of works that appear unusual with respect to the reference corpus. Such investigations can be carried out in a manner that is quantifiable and whose confidence of prediction can, to some extent at least, be evaluated both in terms of the accuracy of its summary, but also in terms of the quality of coverage of the music under discussion the encoded corpus offers. Not only can we quote the accuracy with which our algorithms model observed human behaviour, but we can also observe the amount of data we have for a given area of the corpus, noting, for example, that our observation of trends in Latin-American music may be less reliable than those for classic rock music on the grounds, say, that the proportion of released recordings of the former that are represented in our collection is smaller than that of the latter. The metrics used are less important than the fact that we can begin to make real, meaningful estimates of our uncertainties.

It is important to bear in mind, however, that the evaluation of musical relationships is not a task amenable to automation. The quantification discussed above is a statistical one and, whilst its usefulness will be greater as more information is provided to the system, it is cognitive experiment and musicological reasoning that must prove the final arbiter of the system's performance. Furthermore, such an approach can only offer limited assistance to those wishing to perform detailed analyses of single works—which is the standard paradigm in traditional music analysis. If we follow de la Motte's (1990, p. 8) tri-partition of the tasks of musical analysis for a single musical piece into a) identification of details (which we call features here), b) explanation of how details function as part of a whole and c) choosing a good perspective for presenting a convincing analysis, then corpus-based and computer-assisted musicology can only provide support for the first two tasks. Even identification of features is only possible automatically if each feature, or its components, are known to the system. Although Artificial Intelligence methods can lead to the identification of novel features, many musical features only make sense after the combination of a large number of smaller elements, with intermediate combinations having no meaning, and this makes algorithmic discovery difficult. Appropriate analysis algorithms can basically identify and count features and, by reference to frequency counts in a reference corpus, it is furthermore possible to determine whether the particular combination of features in a given piece is common, relatively infrequent or something in between. But the evaluation of what an infrequent combination of features means in cultural and aesthetic terms is still entirely up to the researcher that makes use of a corpus-based analysis system. In the context of the M^{4}S pop music database, an infrequent combination of features in a given piece can arise from, for example, compositional incompetence where the composer was unable to stay within the boundaries of a particular style, or it can arise from a masterpiece where a composer extended the boundaries of that style to achieve a special aesthetic effect. And of course, an infrequent combination of features can also indicate gross errors in transcription or MIDI encoding. It is important to remember that the interpretation of the (corpus-based) analysis remains the duty and the responsibility of the researcher and cannot be loaded onto the methodology.

The introduction of the term corpus-based musicology is also a deliberate link to the research paradigm of corpus-based linguistics which has seen a tremendous boom in recent times with fast-growing number of electronic text libraries and, of course, the internet as a virtually infinite source of digitally encoded texts. While there are many clear commonalities and differences between the methodologies of statistical natural language processing and statistical music analysis, we would like to mention here only a few of the obvious parallels between the two corpus-based approaches (see, e.g., the introductory chapter of Manning & Schütze, 1999, for the main ideas of the statistical approach in natural language processing). While grammaticality has been a central and binary category for analysing and describing linguistic units such as sentences for a long time it has become apparent in the last decades that the grammaticality of sentences is rather an attribute on a continuous scale which ranges from indisputably wrong to indubitably correct. The continuous nature of the attribute arises from the impact that usage frequencies of linguistic structures have upon human language processing. Language grammars try to fix a set of rules from which grammatical sentences can be derived. But it is easy to come up with examples of both grammatically correct but meaningless sentences or grammatically correct sentences where the correctness is very hard too judge and that are potentially ambiguous and hard to understand. In music, strict obedience to rules of music theory is also not a property through which a piece of music would be described as "wrong" or "right" in modern musicology. The notion that a music style or genre is better characterised by frequent structures, formulae, and note transitions is generally recognised as being more viable than compliance to music theoretic rules.

As musical styles and genres, especially in popular music, change much more quickly than the usage of linguistic structures in natural language, it is generally acknowledged that musical rules-of-thumb and frequent formulae change over time and that this change normally happens gradually. Analogously, the process of change in the usage of linguistic structures, although it generally happens much more slowly, is an established fact. Therefore, it is advisable to choose a corpus with elements (text documents or music pieces respectively) that were created within a limited time-span in which the change in frequency of linguistic or musical structures is negligible or where it can be controlled, e.g. by filtering according to meta-data.

Several of the major goals of linguistic and musical corpus-based analysis overlap to a large degree. This includes the detection of regularities or probabilistic grammars in a corpus, the retrieval of items from a corpus according to a query or search parameters, the grouping and ordering of items within a corpus according to similarity, and last but not least the inference about human cognition in the particular domain.

Finally it is worth mentioning that the statistic tools from corpus-based linguistics have been adopted quite successfully for music analysis in recent years (e.g. Downie, 1999; Conklin & Witten, 1995; Pearce & Wiggins, 2006; Potter, Pearce & Wiggins, 2007). While the basic elements and features (or tokens) over which statistics are computed naturally differ between linguistics and musicology, the statistical concepts that allow us to infer regularities within the specific domain are quite similar or nearly identical. Among the chief statistical concepts that can be derived from frequency counts of tokens/features, and that are employed in both fields, are Markov models, entropy and mutual information, association measures, unsupervised clustering techniques, and supervised classifiers such as decision trees.

3 High level feature descriptors

3.1 Introduction

As those researchers dealing with music information retrieval from audio recordings are made very aware, the music we hear, remember and describe is very hard to discover in the sound itself. Even fundamental concepts such as 'pitch' and 'note' clearly simplify and summarise a complex reality, whether that summary takes place in the ear or in the brain or some combination of the two. These apparently quite basic concepts can then be progressively combined and summarised into other more elaborate constructions, such as interval, melody, sequence, harmony, metre or form. We call all such musically-relevant constructions *musical objects* that can be described by and possibly computed from a set of features.

For our purposes, notes, and their attributes of pitch and rhythm are taken as basic or *atomic*, not because we believe them always to be so, but because in symbolic music, they are not explicitly dependant on any simpler elements, being provided by the representation itself. As we have indicated above, were we using recordings, or, for that matter, graphical scores, this might not be the case. In traditional computer-science style, we describe musical objects that are basic in the above sense as 'low level'. 'High-level' musical objects, then are those that are derived from combination of or deduction from other musical objects.

For conceptual clarity as well as for practical implementation reasons, it is helpful to distinguish between features that summarise emergent features of musical objects, such as a harmonic label, and those that model transformations of a sequence of basic musical events/concepts (i.e. notes) or musical objects. In the former case, the input to a feature computation procedure is a set of musical objects at a given structural level, and the feature, computed as output, is a more abstract set of information. By contrast, in the latter case, the feature is a one-to-one mapping between its input elements and its output elements, in sequence. Examples for these transformational features include accent strength values or entropy values for all notes in a melody notes. The distinction between these two types of features, i.e. summary and transformational features, is useful when correlations or associations between different features are computed and probability distributions of feature combinations are of interest (see description of music corpora as joint probability distributions, above). The toolbox for melodic feature computation, MELFEATURE, designed by Klaus Frieler and Daniel Müllensiefen is implemented to run in two modes, one of which transforms note events of a melody whereas the other one summarises them (Frieler, 2007).

Within the context of M^4S, we focus on some features that we expect to be relevant in connection with the cognitive processing of melodies. The following list contains the summarising features that will play a role in the project along with references to work from the M^4S environment:

- Melodic contour: see below for an in-depth discussion of this feature (Frieler et al., in press)
- Phrase segmentation: Segment a stream of melodic notes into meaningful 'chunks' or phrases (Müllensiefen et al., 2007)
- Harmonic labeling: Summarise the harmonic movement of a time window from a polyphonic piece of music by one or several chord labels in sequence (Rhodes et al., 2007).
- Rhythm classification: Describe the rhythm of a melodic sequence by classifying it into one of a few rhythm classes or by positioning it in a multi-dimensional perceptual rhythm space (Allan et al., 2007)
- Instrumentation/arrangement: Characterise the density and the type of instrumental ensemble playing over a region of a polyphonic music segment (Allan et al., 2007)

Among the transformational features to play a crucial role are:

– Accent structure: Determine the perceptual (accentual) strength of notes of a melody (Pfleiderer & Müllensiefen, 2006).
– Expectedness/Entropy/Information content: Determine the information content/expectedness of each note in a melody (Pearce & Wiggins, 2006).

The literature contains several different versions of each of these features, which differ considerably with respect to how a feature is computed from symbolic music data and what feature value is assigned for a given musical object; the discussion of melodic contour as a feature, below, serves as just one example here. Our approach to the selection and/or construction of features is to implement various alternative and possibly competing versions of the same feature and test, in reliable music-psychological experiments, which versions of a feature best approximate those used by listeners when they cognitively process melodic content. A similar approach has been used successfully by Müllensiefen and Frieler (2004, 2007) for the comparative evaluation of melodic similarity measures.

In addition to the difficulty of choosing from a variety of versions of a particular feature, there are several dangers involved in building a feature model of music pieces in a corpus using high-level features. Since each high-level feature depends on those lower in the feature chain, the reliability of each is compromised by errors in other algorithms. Errors may be compounded to result in increasing inaccuracies, and natural ambiguities, which characterise many musical features, can be lost in the complexity. As an example one can think of melodic phrase contour as a high-level feature which depends on the extraction of a melodic line (i.e. the tune) from a polyphonic piece and subsequently on the segmentation of the notes of the melodic line into melodic phrases. By the time we use these incorrectly-segmented, inaccurately-extracted melodies to find melodic contours, the results may bear no relationship at all to the answers that a listener might produce – in fact they may well appear random. Errors on lower levels can easily propagate onto higher feature levels and may introduce noise in a resulting feature data set.

By comparing multiple algorithms (and parameter choices for each of them) with experimental data, we aim to minimise this sort of error accumulation. Furthermore, the data from the perceptual experiments can help to inform an assessment of the ambiguities in a given task (or even different schools of response to it) and enable an appropriate response.

3.2 Example: Melodic Contour

Melodic contour has a long history as being regarded as one of the most important features in the context of modelling music perception and melodic memory. The contour of a melody is believed to be abstracted early in the listening process and to remain a stable and reliable representation especially for novel melodies. On a very general level contour is conceptualised by most researchers as the pattern of up and down movements of a melody in pitch space over time (see below for several different realisations of this basic concept). Numerous psychological studies in the past have found contour to be a decisive mental representation of melodies or short melodic phrases which memory is based primarily upon, particularly after short retention intervals and for novel or non-standard melodies (for a full discussion of the importance of melodic contour for memory see Müllensiefen, 2004. The long list of melody contour studies includes Dowling & Fujitani (1971), Dowling (1978), Edworthy (1985), and Cutietta & Booth (1996).

Dowling (1978) proposes that the mechanism through which contour works as an effective memory representation can be summarised very briefly and superficially as follows. During listening the up and down movement of the melodic line (=contour) is recorded in memory. Encoding and retaining up and down movements requires the processing of much less information than, say, the retention of the exact size and direction of the melodic intervals and is, therefore, more efficient. At the same time the scale of the melody is abstracted as well. Then, to reconstruct the melody from memory, experimental participants seem to align the retained contour movements to scale degrees from the retained scale and thereby reconstruct melodic intervals. Dowling's scale and contour theory of melodic memory is widely cited throughout the literature and has been vastly influential. But regardless of whether melodic contour is used in exactly the way described here, experimental evidence suggests that contour is a very efficient memory representation.

In the greater part of the psychological literature, melodic contour is defined in a way that allows for a construction of experimental stimuli. Contour is often contrasted with the concept of melodic interval. The former is generally believed to encode only the direction of a melodic interval, or the direction and a gross indication of interval size (e.g. unison, step, skip). While this realisation might be sufficient to separate artificial experimental items in terms of factors in an ANOVA design, it is unsatisfactory from a music-analytical standpoint, and, indeed, musicologists have proposed a number different realisations of contour for analysing existing melodies. Without going into too much detail, we briefly outline five different realisations of contour from the literature that can be regarded as competing with regard to which representation listeners actually use when cognitively processing novel melodies.[2]

Huron Contour

David Huron (1996) proposed a representation of contour based on the pitch height of the first and last notes of a melodic phrase as well as the average pitch of all notes in between. Huron describes phrase contour by the relation the average[3] and the last note, respectively, has with its predecessor. The value of this relation is ordinal which means e.g. that the average note can be higher, lower or equal to the first note but no information about the amount that it is higher than its predecessor is encoded. From the combination of these two free parameters describing relative pitch height on an ordinal level, Huron defines creates distinct contour classes to which any melodic phrase can be assigned. Figure 1 depicts Huron's nine resulting contour classes.

[2] The names of the different contour realisations are used not as such in the literature but are assigned here by the authors to facilitate discussion.

[3] The average note is, of course, only an abstract note

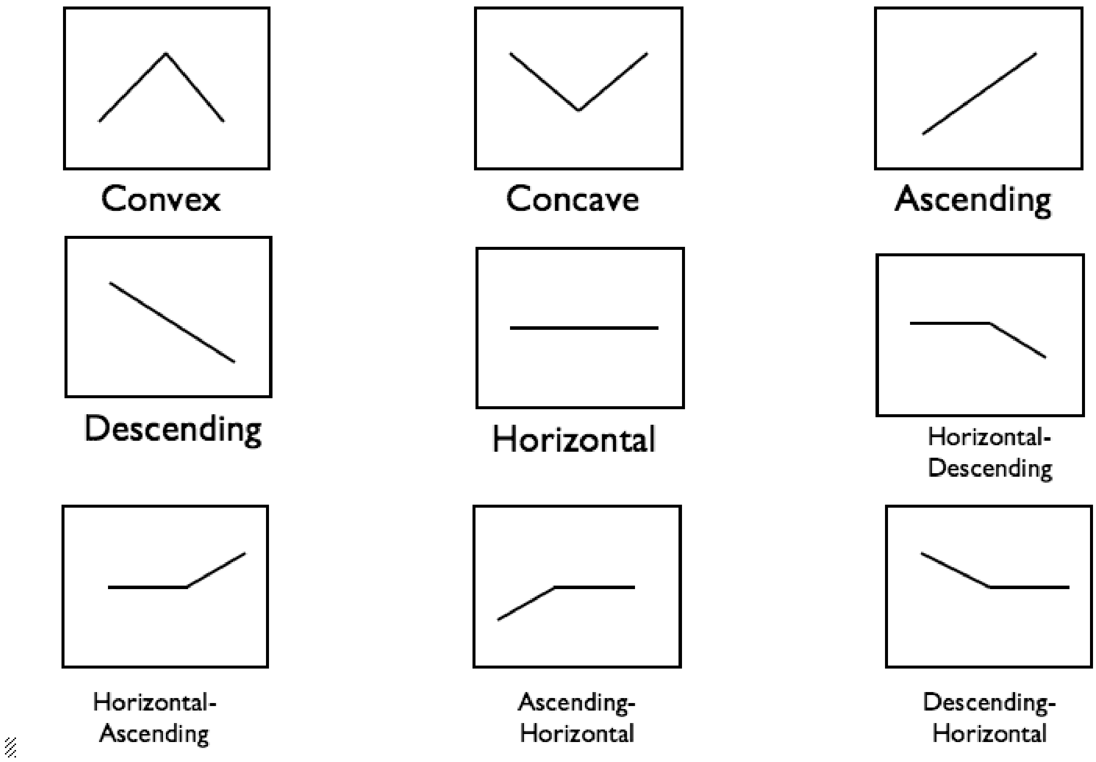


Figure 1: The nine classes of melodic phrase contour according to Huron (1996, p. 8).

The conceptual advantages of Huron's contour model are its computational simplicity and the restriction of the outcome to just nine classes or two free relation parameters. On the other hand, assigning all possible melodic phrases to one of nine contours may be too simple for a describing the cognitive representation of listeners especially if they have a stronger musical background. Another conceptual disadvantage arises from the inclusion of the contour classes which contain a horizontal relationship as one of the free parameters. These classes might have been included in the scheme primarily for the sake of combinatorial completeness but for a lot of melodic phrases found in existing music they might not reflect what was conceptually intended. Imagine the melodic phrase in Figure 2. The starting pitch is C_5 (=72 in MIDI pitch encoding), the last pitch is G_4 (=67) and the average of the intervening notes is (76+74+66)/3 = 72.

Figure 2: Example phrase leading to an assignment of a 'horizontal-descending' Huron contour.

This phrase would be classified as horizontal-descending according to Huron's scheme but it is not impossible to imagine that perceptions might tend rather towards a descending contour shape. It is easy to find similar examples where the inclusion of Huron's hyphenated contour shapes lead to counter-intuitive contour assignments. In fact, it may be more appropriate to define the few musical instances where these hyphenated classes actually do make sense. These instances would for example include a repetition of notes or a trill with a low ending note for the horizontal-descending class.

But for the remainder of this paper we will refer to the original formulation of the Huron contour classes, leaving any modifications or amendments to a future paper.

Extended Huron contour

Huron (1996) also suggests an extension of his nine classes to include 'M' and 'W' shapes. These shapes would be constructed analogously to the contour classes explained above but would comprise 4 free parameters describing ordinal relations instead of 2. Depending on how boundaries of melodic phrases are set in a music corpus—for his explorations of contour frequencies Huron relies mainly on the expert-marked phrase indications in the Essen folk song database—and how many notes a phrase comprises on average, either the extended Huron contour can be constructed as concatenation of two phrase contours or a decision has to be made where to position a middle note which represents the end of the first half and the beginning of the second. This could be either a note near the temporal middle of the phrase, a relatively long note in a middle region, or a contour turning point in a middle region, e.g. selected in such a way that the length of the melodic movement on both sides of the turning point is maximized.

Exploiting all possible combinations of the four 3-valued parameters would result in 3^4=81 possible contour shapes. Even if part of these shapes can be thought of being equivalent (e.g. rising-rising = rising), there still seem to be far too many to be used cognitively efficiently, but one could propose solutions where several similar shapes are mapped onto the same category/class.

Interpolation contour

The idea behind interpolation contour is that melodic turning points are important and salient points in a melody and that notes in-between these turning points can be summarised by a line. The concept of summarising a general melodic movement by a line has been around in music analysis for a long time and was popularised e.g. by Heinrich Schenker's (for examples see Schenker, 1932), *Urlinie*. Wolfram Steinbeck (1982) was one of the first to adapt to the concept for the computer analysis of melodies. An example from his book of how an interpolated line is drawn is given in Figure 3. The beginning of the first note and the end of every contour turning point, as well as the end of the last note, are connected by straight lines in a two-dimensional time-pitch space. A melodic phrase is then characterised by a list of two vectors reflecting the length and the gradient of the interpolation lines. The number of free parameters for this interpolation line contour model is then $2(t+1)$ with t being the number of melodic turning points and the values of the vector elements represent melodic contour on an interval level. This contour representation depends crucially on the method for identifying the melodic turning points in a phrase and differences have been found in a study of melodic similarity where two different realisations of interpolation line contour control for trivial change notes in a different way (see Müllensiefen & Frieler, 2004). Another conceptual difficulty is the variable length of the two vectors which is a function of the number of turning points found in the melodic phrase. This variable length makes comparisons between melodic phrases, that rely on the computation on correlation or difference coefficients of the two vectors, difficult. Zhou and Kankahalli (2003) use a very similar contour representation for a modern query-by-humming system by drawing straight lines ("slopes") between "peaks and valleys" of melodies and then working with the "pitch ranges" and "duration values" of the slopes. This is essentially and equivalent to using the gradient and length of the interpolation lines proposed by Steinbeck (1982) but Zhou and Kankanhalli employ a different terminology and apply this contour concept to a popular task from music information retrieval.

Instead of summarising melody notes by interpolation with a straight line, a curve can also be fitted to the melody notes to be summarised. Fitting a polynomial curve to the notes of a melody was originally proposed by Steinbeck (1982) as well, but due to the lack of computational resources not realised at that time. Frieler et al. (in press) took up that approach and fitted 2nd and 4th order polynomials to 989 melodic phrases taken from commercial pop songs from 1990-2005. Once the order of the polynomial is chosen, the polynomial coefficients can be regarded as a representation of the melodic contour. A comparison of contours of different phrases like in the clustering of polynomials coefficients that Frieler et al. carried out is possible. Using unsupervised model-based clustering via the expectation-maximisation algorithm (see Frayley & Raftery, 1998, for model-based clustering) Frieler et al. found four clusters derived from 2nd order polynomials that were comparably close to a subset of the contour classes proposed by Huron (1996) as Figure 3 shows.

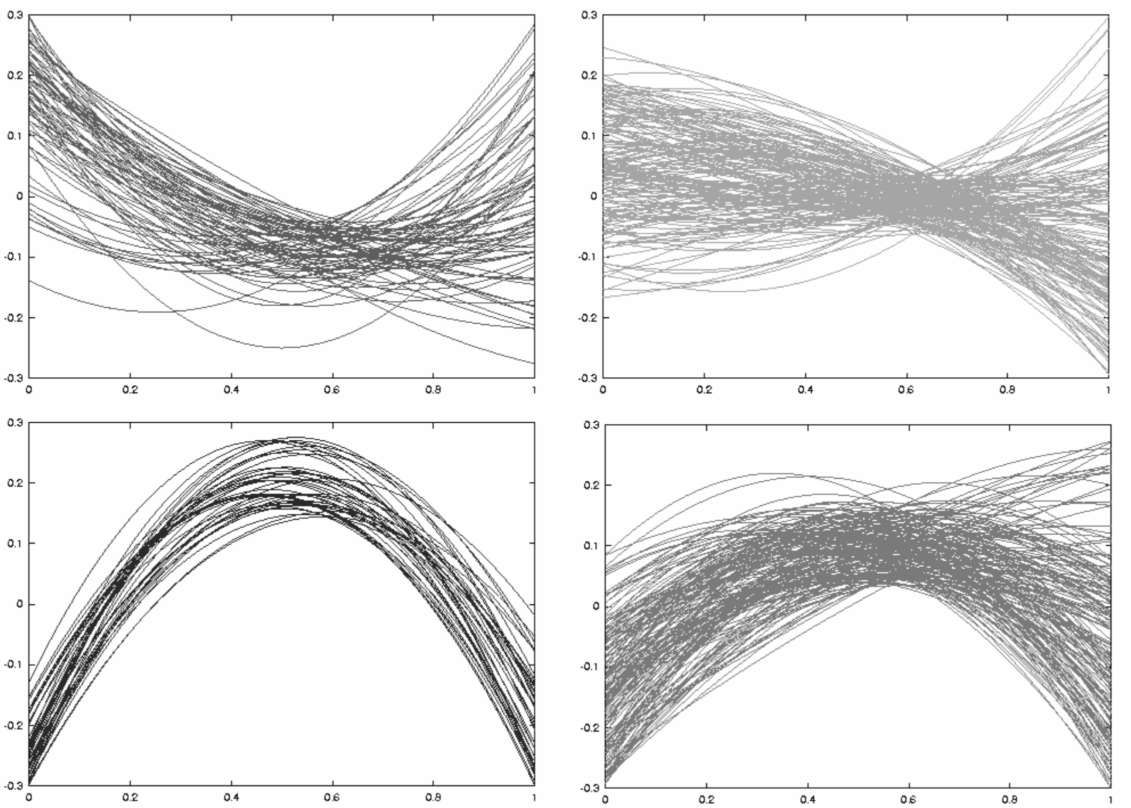


Figure 3: Clusters of interpolation curve contours as found by Frieler et al. (in press). Clockwise from top left: Cluster 1 (concave-descending); Cluster 2 (Mixture); Cluster 3 (Small arch); Cluster 4 (Big arch)

Just as for interpolation line contour, for polynomial curve interpolation contour, the free parameters represent contour on an interval level. But unlike the vectors of length and gradient resulting from the interpolation lines, the set of coefficients of the interpolation polynomial curves is limited to *o+1* where *o* designates the order or the polynomials. So the researcher is free (or rather obliged) to choose the level of observational details for a given research question which might depend on the question to be answered and the musical repertoire at hand. While selecting, too high an order might overfit the melodic phrases and introduce noise into the data, the choice of too low an order might obscure important differences between phrases. In any case, post-processing procedures like principal component analysis might be helpful in reduce reducing noise in the data. Generally, the curve fitting approach to obtaining representations of melodic contour offers a vast set of possibilities which is currently completely unexplored and which includes potentially fitting b-splines, wavelets or procedures using other basis functions to the numerical values of the notes of a melody.

Step curve contour

Representing a melodic contour by a step curve contour can as well be depicted by a graph in time-pitch space where the height of the steps corresponds to pitch height, the beginning corresponds to onset times and length corresponds to the inter-onset interval between two successive notes. The information in this graph can be expressed as a list of two vectors containing length and relative pitch height of all notes in the phrase. This is the richest (= least abstract) contour representation and the only one that allows full reconstruction of the melodic phrase. 2n free parameters are necessary for this representation, where n is the number of notes. Instead of using this compact representation, it is more practical in certain application to sample from pitch values from a step curve at regular intervals (Juhász, 2000; Eerola and Toiviainen. 2004). When melodies are normalised in time, vector correlation and difference measures work straightforwardly for any pair of melodies, as vectors built from sampling necessarily have the same length. Figure 4 shows sample points taken at a sample rate of 4/beat from a step curve contour as presented by Eerola and Toiviainen (2004).

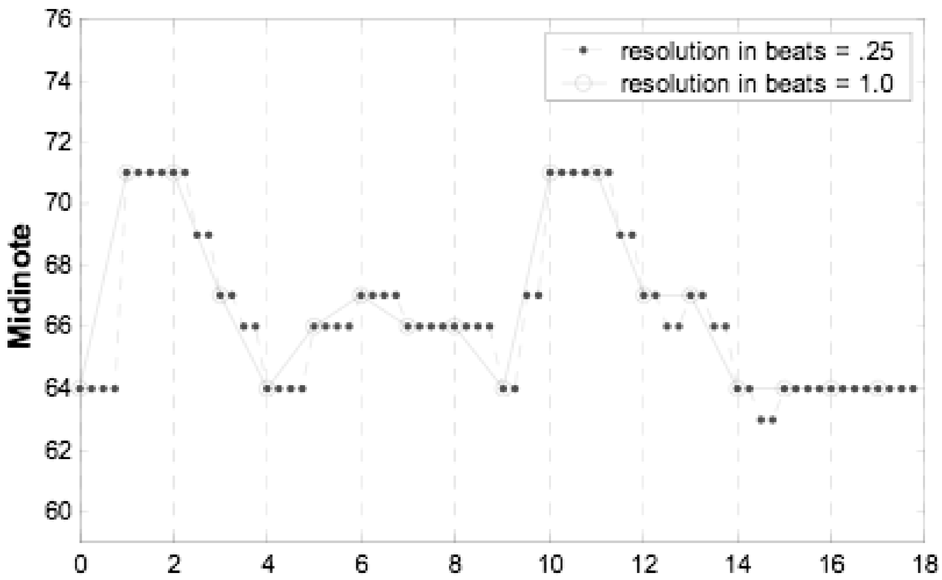


Figure 4: Sampling of melodic contour according to the step curve contour representation. Taken from Eerola & Toiviainen (2004).

The point of our rather detailed discussion of different versions of melodic contour was to show how diverse the formulations and computations of contour in the literature are. With a number of different contour definitions the question for the 'best' contour model or, in our context, rather the model that is cognitively most adequate (i.e. comes closest to the mental contour representation that listeners employ for cognising melodies) becomes very apparent. The way to deal with various competing versions or realisations of a feature within the framework of M^4S is to test it in a rigorous experimental paradigm and with participants from a population which should be as close as possible to the phenomenon to be modelled in the end, i.e. memory for melodies of people with no special music training. As we are unaware of any existing comparative study of different contour definitions, an experiment is current within M^4S which should give indications about which versions of contour are appropriate in the research context of the project. The basic paradigm is a cross-model recognition paradigm where a melody item is presented aurally and should subsequently be identified on a list of visually presented contour shapes. This paradigm exploits the fact that, even for untrained participants, the analogy between auditory movement in pitch space and visual movement on a vertical axis is easily understood and that the fact that all these versions of contour lend themselves naturally to interpretation as visual shapes. The experiment is due to be carried out in early 2008 and its results will be reported in due course.

4 Results

We present in this section some preliminary findings that show several different ways in which the corpus-based approach can generate interesting results. The four types of results presented fall into the categories a) comparison of feature distributions between music corpora, b) identification of generic patterns within a corpus, and c) the identification of specific, non-generic patterns in a corpus. As the M^4S project is an ongoing research enterprise, we will have to spare the application of corpus-based musical information for modelling human music cognition for a later publication.

Comparison of feature distributions between music corpora

If a feature is considered to correspond to a cognitive important entity in the process of music production or perception than comparing the distributions between two corpora can reveal something about why two music styles/corpora are perceived to be different and how large the difference is in quantitative terms.

We chose here melody contour as a feature which is believed to be of certain importance for music production and perception (see discussion above), and we choose Huron contour as an realisation of contour. This is not to mean that we believe Huron contour to be superior to the other contour realisations presented above. As said before the results of the comparative study on the cognitive validity of contour representations are still pending. But Huron contour has the advantage that the computation of the contour classes is relatively easy to understand and that empirical investigations have used this contour feature before (e.g. Huron, 1996).

We counted the frequencies of the nine contour classes for the melodic phrases in the European tunes from Essen folk song database and in the M^4S pop song database. The 5,739 folk songs contained a total of 33,504 melodic phrases. Phrase boundaries are marked up in the Essen collection and were originally generated by the individual transcriber or encoder of the song in question. Phrase boundaries might have been inserted according to the music-analytical understanding of the transcriber/encoder in most instances, but phrase segmentation might have also been carried out according to indications of the song's lyrics in other instances. As no phrase segmentation is provided in the original data of the M^{4}S boundaries, we used the SimpleSegmenter (Müllensiefen & Frieler, 2004) procedure to identify melodic boundaries which in turn provided us with 444,107 melodic phrases from the 14,067 pop songs. The results of the frequency counts are summarised in table 2 and contrasted with frequency counts over 33,504 melody phrases from the Essen folk song collection.

Contour class	Essen folk songs (%, n=33,504)	M4S pop songs (%, n=442,107)
Convex	33.66	23.99
Concave	7.37	11.55
Descending	24.87	16.29
Ascending	16.54	22.85
Horizontal	2.2	5.65
Horizontal-Descending	6.01	5.18
Horizontal-Ascending	1.95	3.53
Descending-Horizontal	2.64	3.7
Ascending-Horizontal	4.76	7.27

The main claim originally made by Huron (1996) on the basis of melodic phrase data from the Essen collection, can obviously be maintained when we look at the M^{4}S pop songs[4]: the melodic arch (i.e. a convex phrase contour) is the dominant model for the melodic movement on phrase level, although the difference in frequency between convex and other contour classes is much less pronounced in pop melodies than for the folk song phrases. This is also corroborated by the high frequency of the two ascending-horizontal and horizontal-descending contours which would form together an arch shape. An interesting difference between the two frequency distributions concerns the prevalence of descending vs. ascending phrases. While we count clearly more descending than ascending phrases in the folk song collection the reverse is true for the pop songs corpus. Also, the horizontal phrases and the partially horizontal contours account for a much larger percentage in the pop song corpus. However, we must be cautious at this stage about basing musicological interpretations on these numbers as we will have to investigate the role of the phrase segmenting algorithm onto subsequent

[4] There are minor differences between our count of the contour classes in the Essen corpus and the number that Huron published (1996) which are due to the fact that we only included European folk songs in our corpus, and that we rounded the average pitch to the nearest semitone. Huron did not round his averages but reasons in a recent personal email communication that "of course, rounding [the average pitch] "makes perfect sense".

contour determinations. Preliminary results from a comparative study on different phrase segmentation models show large differences between different approaches and a suboptimal performance of SimpleSegmenter (Müllensiefen et al., 2007).

Identification of generic patterns

The identification of generic or overly frequent patterns in a corpus can serve two purposes: Firstly, to find the basic building blocks from which longer and more complex musical objects are made from, or phrasing it more poetically, to establish the alphabet that the music of a particular style is written in. Secondly, only once it is know what the generic patterns of a particular corpus are, one is able to judge what a significant or original pattern within the context of a corpus is. We will exemplify this briefly by giving examples from the harmonic and melodic sequences of the M^4S pop database.

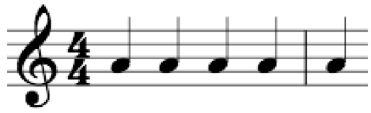
Using the chord labelling algorithm based on Bayesian model selection proposed by Rhodes et al. (2007) we assigned chord labels to all bars of all the 14,067 pieces that had pitched instruments playing. For the current purpose we ignored information about functional bass notes and chord extensions and counted sequences of differing of chord modes (major/minor) and intervals between subsequent chord roots. A preliminary evaluation of the frequency table of harmonic sequences gave the following results:

– 52% of all chord sequences of a maximal length of five subsequent chords are combinations of major chords that are 5, 7 and 2 semitones away from each other, or expressed in terms of a Roman Numeral analysis, we found mostly chord sequences chords in the pop music database that are combinations of I, IV, and V chords.
– A very frequent group of chord sequences can be summarised as combinations I, vi, IV, and V chords. The sequences from this group cover 5% of all sequences in the database and appear at least once in 16% of all 14,067 songs. For the particular sequence of this group where the chords appear in the order given above, several researchers have proposed names like the “doo-wop progression” (Moore, 2006) or the turn-around formula (Kramarz, 2006).
– Very frequently we also find alternations between I and vi or I and ii chords in our pop song database, each which makes for approximately 3.5% of all chord sequences.

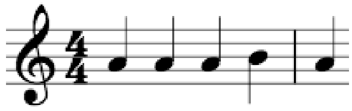
It is difficult to draw a definite line between what can be regarded as generic harmonic patterns and what makes an original harmonic progression but the huge difference in frequency is certainly useful for distinguishing between the two.

For melodic generic patterns we have a very similar picture with the most frequent patterns covering a large percentage of all melodic sequences in the database. For a preliminary evaluation we used only the melodic lines encoded on MIDI channel 4 of the original files which, by a convention used by our MIDI-file provider, contains the main vocal line. We ensured that the main melodic line of each song be a pure monody by an algorithm which adjusts note onset and note end times to avoid small overlaps and, where there are segments of true polyphony, favours louder, higher notes as the main voice. For all monodic lines in the corpus interval sequences of different lengths are counted. Among the most frequent sequences with a length of four intervals (five notes), while ignoring all temporal information, are:

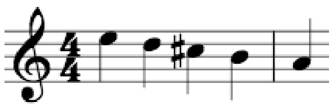
Note repetitions (frequency rank #1)



Combinations of note repetitions and seconds, e.g. rising major second on third and falling major second down on the fourth interval (frequency rank #5)



The first interval sequence with a range of more than a major third appears at frequency rank number 35 and is the descending major scale.



Again it is difficult to define what a generic melodic pattern is but the three very frequent interval patterns we just presented can surely be regarded as the basic building blocks from which pop melodies are made.

Identification of relevant patterns in a corpus

In contrast to the generic patterns just discussed finding special or relevant patterns in a corpus appears to be a much more interesting retrieval task and can be potentially important for various goals, including the detection of cover songs and copyright infringements or for tracing citations and explicit musical references. Again, we have to limit ourselves here to a few examples.

The song *Brown girl in the ring* was at the core of a law suit that spanned almost 25 years in German courts (Pendzich, 2004, p.226). If we represent the six beginning notes of the melody as a sequence of tuples of pitch intervals and duration ratios ((0,1) (0,.75) (-2,.3) (-2,4) (+4,.75)) and query the M^4S database with this sequence we find 26 different songs where this melodic sequence is used. We would like to highlight the comparatively low number of occurrences of this melodic sequence, even though it is a very short and very simple sequence in musical terms. Among the songs that incorporate this melodic sequence are, apart from Boney M's famous version, Rick Astley's *Together forever*, The Beatles' *Maxwell's silver hammer*, and Bon Jovi's *In these arms.* It would be an interesting musicological task to evaluate the significance of this melodic sequence in the respective songs and to investigate whether and where there are any connections between the usage of this formula in Boney M's version or the original Jamaican children's rhyme and occurrence in the other 25 songs.

We carried out an analogous search for the opening harmonic sequence from *Yesterday* by The Beatles which was represented as sequence of chord modes and intervals between subsequent chord roots. In Roman Numerals this sequence can be written as I vii III vi IV. We found only 14 out of 14,067 songs that contained this harmonic sequence. The result set from the database query included apart from the Beatles song and one cover version *Make me smile* by Chicago and *Sara* by Jefferson Starship. Again, note that the number of occurrences of this harmonic sequence is by magnitudes lower compared to the doo-wop progression which we found in more than 2200 songs. Interestingly enough, Mauch et al. (2007) report a much stronger prevalence of this sequence in a corpus of Jazz pieces from the Real Book. Again, with the support of these quantitative findings an interesting socio-cultural explanation might connect the success of that chord sequence to the Beatles' interest in Jazz harmony at a certain time.

5 Summary and conclusion

The aim of this contribution has been to present a modern approach towards modelling human music cognition. This methodology is employed in the current M^{4}S project on melody cognition and memory processing for melodies. The main frameworks that support this approach are the automatic extraction of cognitively relevant features from symbolically encoded music and the statistical analysis of a large music corpus which is annotated in terms of the relevant features. The idea is to model the implicit knowledge of western pop music listeners by the proxy of a statistical description of the music corpus that the listeners build their listening experience on. The assumption is that using information about the prevalence of musical features in real music within the modelling framework is substantially better than merely applying musical features to relate music data to behavioural data as if a long enculturation in a musical culture had never happened or was irrelevant.

As the M^{4}S project is a still ongoing research work, we have not presented any results from the memory experiments that are due towards the end of the project to support our claims about the usefulness in cognitive modelling. Instead, we have focused on exemplifying the methodology by discussing in depth the feature of melodic contour and by outlining the concept of corpus-based musicology. Although grounded here in music cognition research, corpus-based musicology has potential as a very fruitful approach for a range of areas in music research. We provide a few examples of the type of results a corpus-based approach can generate, although, the true potential of this approach will only become apparent over time as researchers increasingly inform their methods with empirical and quantified knowledge about the corpus of music they are studying.

References

Allan, H., Müllensiefen, D., & Wiggins, G. (2007). Methodological considerations in studies of musical similarity. In *Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 473–478).

Brook, B. (1970). The plaine and easie code. In B. Brook (Ed.), *Musicology and the computer: Three Symposia.* New York.

Clarke, E. & Cook, N. (Eds.). (2004). *Empirical Musicology: Aims, Methods, Prospects*. Oxford: Oxford University Press.

Conklin, D. & Witten, I. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Cutietta, R. A. & Booth, G. D. (1996). The influence of metre, mode, interval type und contour in repeated melodic free-recall. *Psychology of Music*, *18*, 45–59.

de la Motte, D. (1990). *Musikalische Analyse* (6. ed.). Bärenreiter.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354.

Dowling, W. J. & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2, Part 2), 524–531.

Downie, J. S. (2003). *Evaluating a simple approach to music information retrieval. Evaluating a simple approach to music information retrieval. Conceiving melodic n-grams as text.* Doctoral dissertation, Faculty of Information and Media Studies, University of Western Ontario, London (Ontario), Canada. Available from http://people.lis.uiuc.edu/~jdownie/mir_papers/thesis_missing_some_music_figs.pdf

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, *2*(3), 375–388.

Eerola, T. & Toiviainen, P. (2004). Mir in matlab: The midi toolbox. In *Proceedings of the 5th International Conference on Music Information Retrieval*. Available from http://ismir2004.ismir.net/proceedings/p004-page-22-paper193.pdf

Fraley, C. & Raftery, A. (1998). How many clusters? which clustering method? - answers via model-based cluster analysis. *Computer Journal*, *41*, 578–588.

Frieler, K. (2007). Melodic feature machines: Melfeature & melex. Presentation held at Symposium on MIR, University of Utrecht. Available from http://www.cs.uu.nl/research/projects/witchcraft/mir2007/frieler.pdf

Frieler, K., Müllensiefen, D., & Riedemann, F. (in press). *Statistical search for melodic prototypes*. Staatliches Institut für Musikforschung.

Hewlett, W. B. (1997). Musedata: Multipurpose representation. In E. Selfridge-Field (Ed.), *Beyond MIDI: The handbook of musical codes* chapter 27, (pp. 402–447). Cambridge, MA: MIT Press.

Huron, D. The new empiricism: Systematic musicology in a postmodern age. lecture 3 from the 1999 Ernest Bloch lectures. Available from http://musiccog.ohio-state.edu/Music220/Bloch.lectures/3.Methodology.html

Huron, D. (1988). Error categories, detection, and reduction in a musical database. *Computers and the Humanities*, *22*, 253–264.

Huron, D. (1995). *The Humdrum Toolkit: Reference Manual*. Menlo Park, California: Center for Computer Assisted Research in the Humanities.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

International MIDI Association (1988). *Standard MIDI Files 1.0*. Los Angeles: International MIDI Association.

Juhász, Z. (2000). A model of variation in the music of a hungarian ethnic group. *Journal of New Music Research*, *29*(2), 159–172.

Karbusicky, V. (1979). *Systematische Musikwissenschaft*. Fink.

Karbusicky, V. & Schneider, A. (1980). Zur Grundlegung der Systematischen Musikwissenschaft. *Acta Musicologica*, *52*(2), 87–101.

Kramarz, V. (2006). *Die Popformeln*. Voggenreiter.

Lockwood, L. (1970). A stylistic investigation of the masses of josquin desprez with the aid of the computer: a progress report. In B. Brook (Ed.), *Musicology and the computer: Musicology 1966-2000: a Practical Program*. City University New York.

Manning, C. D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Mauch, M., Dixon, S., Harte, C., Casey, M., & Fields, B. (2007). Discovering chord idioms through beatles and real book songs. In *Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 255–258).

McKay, C. (2006). jsymbolic: A feature extractor for midi files. In *Proceedings of the International Computer Music Conference* (pp. 302–305).

Moore, A. (1997). *The Beatles, Sgt. Pepper's Lonely Hearts Club Band*. Cambridge: Cambridge University Press.

Moore, A. (2006). What story should a history of popular music tell? *Popular Music History*, *1*(3), 329–338.

Müllensiefen, D. & Frieler, K. (2004). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology*, *13*, 147–176.

Müllensiefen, D. & Frieler, K. (2007). Modelling experts' notions of melodic similarity. *Musicae Scientiae*, *Discussion Forum 4A*.

Müllensiefen, D., Pearce, M., Wiggins, G., & Frieler, K. (2007). Segmenting pop melodies: a model comparison approach. Presentation held at SMPC07, Montreal, Canada.

Parncutt, R. (2007). Systematic musicology and the history and future of western musical scholarship. *Journal of Interdisciplinary Music Studies*, *1*, 1–32.

Pearce, M. T. & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*(5), 377–405.

Pendzich, M. (2004). *Von der Coverversion zum Hit-Recycling*. LIT.

Pfleiderer, M. & Müllensiefen, D. (2006). The perception of accents in pop music melodies. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th International Conference of Music Perception and Cognition (ICMPC9)* (pp. 1272–1280).

Potter, K., Wiggins, G., & Pearce, M. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Scientiae*, *11*(2), 295–322.

Recordare (2004). Music xml definition 2.0. Website. Available from http://www.musicxml.org/xml.html.

Rhodes, C., Lewis, D., & Müllensiefen, D. (2007). Bayesian model selection for harmonic labelling. In *Program and Summaries of the First Internatinal Confernece of the Society for Mathematics and Computation in Music, Berlin, May, 18-20, 2007*.

Rosen, C. (1976). *The classical style: Haydn, Mozart, Beethoven*. London: Faber.

Sagrillo, D. (1999). *Melodiegestalten im luxemburgischen Volkslied: Zur Anwendung computergestützter Verfahren bei der Klassifikation von Volksliedabschnitten*. Bonn: Holos.

Schaffrath, H. (1995). *The Essen Folksong Collection in the Kern Format*. Center for Computer Assisted research in the Humanities.

Schenker, H. (1932). *Fünf Urlinie-Tafeln*. Wien: Universal Edition.

Schneider, A. (1993). Systematische musikwissenschaft: Traditionen, Ansätze, Aufgaben. *Systematische Musikwissenschaft*, *1/2*, 145–180.

Selfridge-Field, E. (1997). Describing musical information. In E. Selfridge-Field (Ed.), *Beyond MIDI: The handbook of musical codes*. Cambridge, MA: MIT Press.

Smaill, A., Wiggins, G., & Harris, M. (1993). Hierarchical music representation for analysis and composition. *Computers and the humanities*, *27*, 7–17.

Steinbeck, W. (1982). *Struktur und Ähnlichkeit: Methoden automatisierter Melodieanalyse*. Kassel: Bärenreiter.

Zhou, Y. & Kankanhalli, M. S. (2003). Melody alignment and similarity metric for content-based music retrieval. In *Proceddings of SPIE-IS&T Electronic Imaging*, volume 5021 (pp. 112–121).